Appendix XVII.

"Review of Atrazine PRA"

Prepared by:

Douglas Urban, Edward Odenkirchen,
Mary Frankenberry, Jim Lin, William Rabert

U.S. Environmental Protection Agency
Office of Pesticide Programs
Environmental Fate and Effects Division
Ariel Rios Building (7507C)
1200 Pennsylvania, Avenue, N.W.
Washington, D.C. 20460

With Technical Assistance from:

Philip Goodrum Bill Brattin, Pat Durkin

Syracuse Research Corporation
Environmental Science Center
6225 Running Ridge Road
North Syracuse, NY 13212
http://esc.syrres.com

December 6, 2001

PREFACE

This document was developed by the Environmental Fate and Effects Division (EFED) in EPA's Office of Pesticide Program's (OPP) with technical assistance provided by Syracuse Research Corporation (SRC). During September / October 2001, a team of scientists in EFED reviewed the report entitled, "Aquatic Ecological Risk Assessment of Atrazine - a Tiered Approach", dated June 30, 2000 (Giddings et al., 2000). Concurrently, SRC was contracted to critically review the report, and prepare materials for an EFED workshop to discuss EFED staff scientist's and the contractor's perspective on the atrazine assessment. Prior to the workshop, twenty-two copies of the Syngenta's (formerly Novartis) refined assessment of atrazine were distributed to EFED staff for thorough review and comment. The workshop took place over two and one-half days, from October 23 - 25, 2001, and was attended by 19 EFED staff scientists, and representatives from the Special Review and Registration Division, the Health Effects Division, and the Office of Water. The objectives of the workshop were to focus on the four levels of refinement portrayed in the assessment, the assumptions behind each refinement, the decisions made to transition to higher levels of refinement; and to compare the approach used in the atrazine refined assessment to that proposed by EFED's PRA implementation team. Comments from both EPA risk assessors and managers were integrated with contractor input throughout the workshop. Following the workshop, SRC compiled the comments from the workshop into a draft "Review of Atrazine PRA", and submitted it to EFED. Subsequently, EFED modified the draft and developed this document.

RESPONSE TO ATRAZINE PRA

BACKGROUND

U.S. EPA has reviewed the report entitled, “Aquatic Ecological Risk Assessment of Atrazine - a Tiered Approach”, dated June 30, 2000 (Giddings et al., 2000), along with supporting documentation. The document is hereafter referred to as a probabilistic risk assessment (PRA) document / report for atrazine. Specific comments are organized in the same general outline of the report - Problem Formulation, Exposure Characterization, Effects Characterization, and Risk Characterization - for each of four Tiers. In addition, a set of general comments are provided, which reflect an overall pattern or theme that became apparent after reviewing the entire document. Given that the PRA appears to be designed as a case study of the tiered process recommended by the Ecological Committee on FIFRA Risk Assessment Methods (ECOFRAM), the comments reflect an evaluation of this goal. It is important to note that the tiered process and methods developed by ECOFRAM were recommendations to the Agency. Subsequent to those recommendations, the EPA has developed and proposed an implementation plan for PRAs [See http://www.epa.gov/scipoly/sap/2000/index.htm#april , FIFRA Scientific Advisory Panel Meeting, April 5-7, 2000:Implementing Probabilistic Ecological Assessments: A Consultation; and, http://www.epa.gov/scipoly/sap/index.htm#march March 13 - 16, 2001: A Case Study: Advancing Ecological Risk Assessment Methods in the EPA, Office of Pesticide Programs], which while based on the ECOFRAM recommendations, contain methods and approaches which built upon those recommendations. Where differences in approach exist, EPA recommends that the more recent methodologies developed by EFED’s implementation team and supported by the SAP be followed. While the methods and approaches used in the PRA are relatively consistent with those initially identified by ECOFRAM, EFED has recognized limitations in the approach and has further refined methodologies to be more reflective of EPA’s need for greater transparency and conservancy in underlying assumptions. Methodologies that differ from the EPA implementation plan are noted in the following comments.

GENERAL COMMENTS

1. Transparency

A PRA report should present sufficient information that all calculations and results can be replicated. Overall, while the PRA for atrazine was generally well written, it lacks the transparency needed to support the conclusions at each Tier. The following general examples highlight EPA’s concerns regarding transparency:

- insufficient rationale explaining what selected data or literature sources were excluded from the risk assessment. What criteria were used to determine that specific references were not relevant? Were the criteria applied consistently across tiers?
- calculations cannot be reproduced because input assumptions or data are incompletely documented.

2. Sensitivity / Uncertainty Analysis

A sensitivity analysis provides a quantitative measure of the relationship between risk estimates and inputs to the models used to assess exposure and effects. Sensitivity analysis can also be used to explore model uncertainty when alternative choices of modeling approaches and assumptions are available. Sensitivity analysis is a critical component to quantitative uncertainty analysis because it adds perspective to the importance of different sources of uncertainty in risk estimates. As such, it plays an important role in the Tiered process and should be conducted in each tier of analysis. This is particularly important because, with the use of PRA, the tools available and the value added from sensitivity analysis are greatly increased. ECOFRAM (Chapter 3, p. 56) highlights two references that present an evaluation of the relative significance of specific combinations of pesticide and environmental variables (Final Report for the FIFRA Environmental Model Validation Task Force, 1999 [www.femvtf.com]; Fontaine et al., 1992[1]). More recently, EFED has conducted a sensitivity analysis of PRZM3.0 (Wolt et al., 2001[2]). The atrazine document does not report results of a sensitivity analysis, and it is unclear if this was even a step taken in the tiered approach. The following general information should be included in the risk assessment:

- Criteria used to determine that specific input variables should be explored further (or alternatively, are sufficiently quantified given the available information and modeling results).
- Results from multiple simulations with alternative modeling approaches or assumptions, such as the use of regression equations, or probability weighting schemes.
- The effect of applying exclusion criteria to a data set. Specifically, results of the analysis before and after applying the criteria, or an explanation as to why this step was skipped.

---

[1]Fontaine, D.D., P.L. Havens, G.E. Blau, and P.M. Tillotson. 1992. The role of sensitivity analysis in groundwater risk modeling for pesticides. *Weed Technology*. 6: 716-724.

[2]Wolt, J., S. Piyush, S. Cryer, J. Lin. 2001. Sensitivity Analysis for Validating Expert Opinion as to “Ideal” Data Set Criteria for Transport Modeling. Draft prepared for the ET&C Annual Review Edition.

3. Tiered Process

Ideally, by applying a tiered process to risk assessment, there will be an appropriate balance between the complexity of the analysis and the level of protectiveness required to make risk management decisions. Figure 1 provides a conceptual overview of this objective by moving diagonally with each consecutive tier. EPA recognizes that the authors intended the atrazine document to serve as a general "case study" that would illustrate the basic concepts recommended by ECOFRAM. However, EPA believes that some of the decisions regarding the analyses included in a specific Tier, and the data employed in the analysis, are inappropriate. Of specific concern is when a more complex approach is adopted without clearly adding to the quantitative uncertainty analysis of the risks (i.e., a lateral rather than diagonal move in Figure 1).

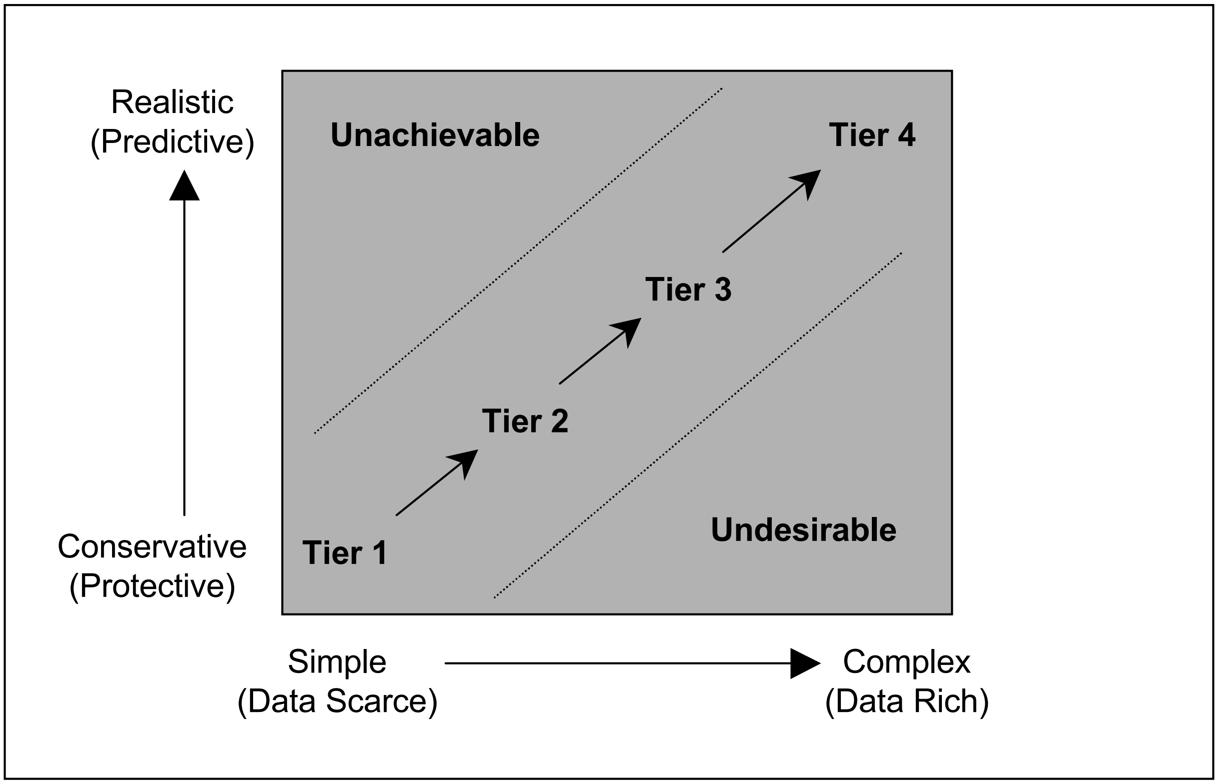

Figure 1. Conceptual illustration of the goal of the tiered process. As the risk assessment progresses to higher tiers, the analysis and modeling approaches may become more complex, while at the same time yielding a more realistic estimate of risk. Risk estimates may increase or decrease with each consecutive tier. The value added for risk managers is measured by the increased understanding of the sources and relative magnitude of uncertainties in the exposure and effects characterization. In the atrazine PRA, there are several examples of “lateral moves” in the diagram, in which complexity was added without informing the quantitative uncertainty analysis of risk.

Another objective of the tiered process is to clearly link the results of a lower tier to the areas of refinement for subsequent analyses. For example, if the problem formulation includes an assessment endpoint such as long-term viability of fish populations, then the risk characterization at each tier should address uncertainties associated with the measurement endpoints (species selected, interpretation of dose-response data and/or toxicity reference values, etc). Does the selection of a particular focal species suggest that it is generally representative of fish across a wide range of aquatic environments for a particular crop/soil/climate scenario? These types of assumptions need to be clearly documented, along with sources of uncertainty introduced by the assumptions. The PRA did not consistently relate the results of each tier to the original problem formulation.

4. Model Availability and Documentation

Every model employed in the analysis should be made available to EPA for consideration. EPA cannot evaluate results that are based on new models, even if the modeling approaches run existing models iteratively, or are described in the risk assessment. At a minimum, copies of the executables, along with a user's guide, and examples of inputs and outputs should be provided - ideally, before the risk assessment is submitted.

5. Selected Assessment Endpoints and the Linkage of Analysis Methods and Results to Them

On page 49 of the risk assessment the following assessment endpoints which purportedly address "integrity of ecosystem structure and function":

- primary productivity;
- sustainability of aquatic macrophyte community structure;
- long-term viability of fish populations.

EPA has concerns that the selected assessment endpoints are somewhat limited in scope. The assessment correctly assumes primary community and macrophyte community structure as important aspects of the analysis, but makes a leap of "functionality faith" to fish population viability without including other potentially important aspects of aquatic systems. The endpoints overlooked may also include structure and productivity of invertebrate communities reliant upon primary producer community structure and productivity.

The three earliest tiers rely on laboratory measurement endpoints for characterization of the above assessment endpoints. However, the assessment is not clear just how the early tier methods and results are to be interpreted in the context of the assessment endpoints. For example, how are species sensitivity distributions related to long-term viability of fish populations in any manner other than measure of direct toxic effects. The absence of such linkages between assessment and measurement endpoints calls into question the validity of assessment conclusions and the direction in which subsequent tiers progress. For example, without analysis of indirect effects on invertebrates or fish, one must question the Tier 3 conclusion that animals are not impacted by atrazine.

SPECIFIC COMMENTS

I. TIER 1

1.A. Problem Formulation

The PRA (p. 49) indicates that with each consecutive tier, measurement endpoints differed while assessment endpoints remained the same. The following assessment endpoints were selected to address the integrity of the ecosystem structure and function: a) primary productivity ; b) sustainability of aquatic macrophyte community structure; and c) long-term viability of fish populations. In Tier 1, the measurement endpoint was defined as acute and chronic risk quotients based on conservative assumptions regarding exposure and toxicity (p. 101). In addition, an objective for Tier 1 was to focus efforts in higher tiers on sensitive taxa and high-risk exposure scenarios. These objectives were generally met based on the exposure and toxicity characterization, although some of the decisions regarding the data selection and analysis are not well supported (see below).

1.B. Exposure Characterization

The scenarios and models used to estimate concentrations of atrazine in farm ponds were generally appropriate and well documented. The goal of evaluating a single high exposure environment representative of each crop / use pattern is appropriate for Tier 1. This includes using maximum applications and conservative parameter values in GENEEC v.1.3 simulations, in order to obtain peak instantaneous (daily) estimated environmental concentrations (EEC).

a. Application of Statistical Analysis without Discussion of Uncertainty

GENEEC calculations were easily reproduced from the parameter values summarized in Table 3.3 of the PRA (p. 111). Input assumptions for GENEEC did appear to be conservative, although there is some concern about applying the following equation for determining the 90% confidence interval for the aerobic soil metabolism $T_{1/2}$ with only two values (p. 105):

$$CI_{1-\alpha} = \bar{x} + t_{1-\alpha,n-1} \frac{s}{\sqrt{n}}$$

where x bar is the sample arithmetic mean (an estimate of μ), *s* is the sample standard deviation (an estimate of *σ*), α is 0.10 (for 1-sided 90% CI), $t_{(1-\alpha,\, n-1)}$ is the t statistic that cuts off (100 α)% of the upper tail of the t-distribution with n-1 degrees of freedom, and *n* is the sample size of data drawn from a normal distribution. Not surprisingly, when applied to such a small sample size (n=2; 21 days and 146 days), the 90% CI exceeded the maximum of the small data set (i.e., 202 days). The PRA appropriately states that as the sample size increases, the 90% CI will approach the mean. If a particular statistical approach is used to demonstrate one method from the “toolbox” of approaches given by EPA guidance (e.g., EPA, 1995), then it is important to underscore the uncertainty associated with the application (in this case, n=2 does not allow for any exploration of the underlying distribution, or even non-parametric approaches). Such a discussion could have been included in the PRA as well as in the useful summary of uncertainties by Tier given by Williams et al.[3] (2000; Table 1.1, p. 87).

Recommendation:

Limitations in approaches used to quantify the conservative input assumptions should be highlighted in a discussion of uncertainty. Methods for calculating upper confidence intervals should be applied thoughtfully, rather than as part of a cookbook approach. The use of “core” data is appropriate, but raises a question as to the suitability of other non-core data at this early tier of analysis. Further discussion and rationale would typically be necessary to exclude other available data, especially if including these data may have yielded even more conservative estimates of exposure.

1.C. Effects Characterization

a. Selective Use of Data

Although EPA recognizes that the authors intended the atrazine document to serve as a general "case study" that would illustrate the basic tiered process recommended by ECOFRAM, EPA believes that in the case of atrazine (where there are extensive toxicity data available), the selective use of toxicity data to "simulate" a typical Tier 1 scenario raised many questions with respect to the specific data selected. For example, how would the magnitude of the Risk Quotients (RQ) derived in Tier 1

---

[3]Williams, W.M., R.F. Pugsley, N.J. Snyder, K. Balu, S.D. Andrish, and W. Chen. 2000. Exposure Assessment of Atrazine in Surface Waters: a Tiered Probabilistic Modeling Approach. Waterborne Environmental, Inc. Project No. 376.01. December 8.

(see Tables 3.5 and 3.6) have changed if the full data set (Appendix B) had been used to select the most sensitive receptors, instead of just the data set that was employed (Table 3.1)? It might appear that this is not a critical issue since a decision was made to proceed to Tier 2 for both plant and animal receptors, but the choice of data for Tier 1 is actually quite important since the same set of receptors was retained in Tier 2 as in Tier 1. Thus, the selective use of data impacted the credibility of both Tiers 1 and 2.

Recommendation:

In general, EPA recommends that all readily available standard toxicity studies that meet normal study quality requirements should be utilized at every tier. With regard to Tier 1, the lowest TRV (Toxicity Reference Value) value for each category of receptors should be used in the calculation of RQ values.

b. Lack of Transparency in TRV Derivation

The PRA indicates that all of the toxicity data that were utilized in all of the tiers of the assessment are summarized in Appendix B. However, it is not clear how the toxicity values used in Tier 1 were derived from the data in Appendix B. For example, in Table 3.1, the 14-day EC50 for *Lemna gibba* (duckweed) is reported to be 43 μg/L. The specific study used to derive this value is not reported in Table 3.1 (it should be). Based on inspection of Table B-2, it seems the most likely study used to derive the value in Table 3.1 was the 14-day study by Hoberg (1993a). However, the data in Table B-2 for Hoberg (1993a) do not include a value of 43 μg/L, and the lowest value reported is 22 μg/L (based on an endpoint of frond production). In accord with the principles identified in ECOFRAM, it appears that the lowest value should have been used at this stage in the tiered process.

Likewise, Table 3.1 lists an acute 96-hr LC50 of 4,500 μg/L for *Oncorhynchus mykiss* (rainbow trout), apparently based on the study of Bathe et al. (1975a). However, another study by Bathe et al. (1976) lists a range of LC50 values from 3,500-5,700, and the value of 3,500 is indicated as having been used in the "acute effects analysis" for freshwater fish. Thus, it is not apparent why the lower value of 3,500 μg/L was not selected for use in Tier 1.

Recommendation:

The document should provide sufficient detail so that the basis of each TRV is clear and reproducible. This includes specifying the study employed, and discussing the basis for selecting the TRV from within the study. If for any reason the Tier 1 TRV selected for a receptor group is not the lowest reported value across all studies for that group, the document should provide a clear rationale and justification for the alternative TRV selected.

c. Evaluation of Taxonomic Groups

Table 3-1 presents the acute and chronic toxicity data utilized in Tier 1. This includes several different groups of receptors, including macrophytes, algae, invertebrates, and fish. However, for the purposes of the Tier 1 evaluation, these were collapsed into just two groups: "animals" and "plants". This approach has the advantage that it simplifies the Tier 1 risk characterization to just four "bins" (acute and chronic plants and animals), but information is lost on whether there is a need to progress to higher tiers for all members of each group. For example, granted that the risk characterization finds that "animals" are of potential chronic concern, does this apply to fish or just invertebrates?

Recommendation:

Major taxonomic groups should be maintained and evaluated separately (assuming the data warrant), so that it is possible to focus more effort on the groups of interest.

1.D. Risk Characterization

a. Inadequate Discussion of Uncertainty in Measurement Endpoint

The RQ approach used for risk characterization in Tier 1 is clear and easily reproducible, and is in accord with the approach recommended by ECOFRAM. However, the risk characterization section contains only limited discussion of the potential uncertainties in the characterization. In particular, there is inadequate discussion of the possibility that the TRVs employed may not be representative of the most sensitive receptors. In this regard, the text (e.g., p. 46, middle paragraph) asserts that use of the LOC values recommended by EPA accounts for the uncertainty in this area, but that is not fully correct. The LOC values recommended by EPA are intended only to account for the extrapolation from a frank effect level (LC50, EC50) to an estimate of the no-effect level within the specific test organism being studied. The LOC does not include an uncertainty factor intended to account for potential occurrence of more sensitive species.

Recommendation:

EPA recommends that the risk characterization section of Tier 1 (and every other tier as well) contain a thoughtful and balanced discussion of the uncertainties and limitations in the risk characterization, including not only those assumptions that are likely to lead to an overestimate of hazard, but also those that might lead to an under-assessment of hazard.

TIER 2

2.A. Problem Formulation

The goals of Tier 2 are well presented:

"Tier 2, like Tier 1, emphasized sensitive species and severe exposure scenarios. However, unlike Tier 1, results of the Tier 2 exposure analysis are expressed as distribution functions. The distributions represented extreme concentrations [1 in 10 year return concentrations] over a range of geographic locations and crop uses." - Williams et al., 2000 (p. 25)

"The objective of Tier 2 was to estimate the probability and magnitude of effects on sensitive species under exposure scenarios that accounted for more of the specific properties of atrazine and incorporated more realistic assumptions about atrazine use and environmental conditions." PRA report (p. 115).

This approach is consistent with ECOFRAM recommendations and EPA's intent, i.e., to allow flexibility to use multiple PRZM/EXAMS simulations to develop improved exposure estimates with multiple headwater systems in multiple regions, as necessary and appropriate to the specific assessment. It should also be understood that Tier 2 analyses will generally continue to focus on conservative exposure scenarios and sensitive species. Conceptual models should characterize temporal variations based on the historical weather patterns for a specific area, and provide reasonably conservative model estimates of pesticide loadings to the standard farm pond. In addition, EPA believes that the resulting EECs should continue to reflect crop/soil/climate conditions that yield reasonable worst case scenarios over multiple years. For example, scenarios should include soils with relatively high runoff and low infiltration potential (i.e., USDA hydrologic soil classes C or D).

The PRA deviates from this conceptual approach by introducing a more refined level of analysis that includes a broader range of environmental conditions, and applies an area weighting scheme to the resulting EECs. While this may be useful information that will help to quantify uncertainty in the corresponding risk estimates, incorporating the full range of conditions in the risk characterization will tend to obscure potentially high-risk conditions, which are still the focus in the earlier tiers. More concerted efforts should be made to identify combinations of environmental factors that would tend to yield higher EEC's. This may involve post-processing the MUSCRAT simulations to perform sensitivity analyses. The quantitative methods that attempt to incorporate the variability in environmental conditions into risk characterization are better suited to a Tier 3 level of refinement to the PRA, rather than a Tier 2 assessment.

Regarding the use of MUSCRAT, EPA has previously suggested (Technical Progress Report of the Implementation Plan for Probabilistic Ecological Assessments: Aquatic Systems, March 15, 2000) that the use of MUSCRAT should be avoided in Tier 2, due to documentation and validation issues, as MUSCRAT is a relatively new tool.

The introduction of distribution functions to represent the toxic response over a range of concentrations is appropriate for Tier 2.

Recommendation:

In Tier 2, exposure assessments should evaluate a range of crop/soil/climate conditions that would be expected to yield more conservative estimates of EEC.

2.B. Exposure Characterization

PRZM / EXAMS simulations were run for standard pond scenarios with specific crops grown in specific regions. The model simulations yielded a 36-year record of estimated atrazine concentrations in pond water averaged over different exposure durations (daily, 96-hour, etc.). A 10-year return period was defined by the 90th percentile of the 36 annual maximum values for each exposure duration. The exposure characterization to this point is clear and easily reproducible, and is in accord with the approach recommended by ECOFRAM; namely, to develop an improved exposure approach with multiple regional water body types (e.g., across 11 regions).

The PRA utilizes the Multiple Scenario Risk Assessment Tool (MUSCRAT) to facilitate a batch run and statistical analysis of exposure scenarios evaluated with PRZM/EXAMS. Again, a fully functional, well-documented version of the MUSCRAT model should have been provided to EPA along with (or prior to) the presentation of the PRA document. Although EPA did participate in the development of MUSCRAT, EPA cannot be expected to accept model results without having the opportunity to independently reproduce the results with a verified (working) version of the model.

The exposure scenarios are designed to explore a variety of different environmental conditions that could affect the concentrations of atrazine in a sensitive headwater farm pond. At the Tier 2 level of analysis, it is understood that some of the conservative point estimate assumptions would be more completely quantified by evaluating a plausible range of conditions. Factors that were considered in the PRA included crop type, soil, climate, and categories of “crop suitability” as defined by U.S. Department of Agriculture (pp. 116-117). Factors associated with the landscape of a particular geographic area (e.g., field slope/length, fraction of treated fields, composition of untreated areas) were appropriately excluded from consideration. The concept of a “bin” is introduced in order to represent runoff/erosion potential, “determined from annual water runoff and sediment yield predicted by 30-year model simulations for each soil/weather combination.” As described, this appears to be a useful, systematic approach for dividing regions into areas of distinct soil/weather conditions. Additional information should have been provided to more fully summarize the differences in characteristics of bins within selected regions.

Recommendation:

When the exposure unit for a simulation changes by subdividing geographic areas into smaller areas according to common environmental properties (e.g., crop use, soil type, climate, etc.), these properties should be summarized in tabular and/or graphic format (i.e., maps). As previously noted, EPA does not recommend the use of MUSCRAT for Tier 2 at this time.

a. Application of Advanced Modeling Assumptions without Sensitivity Analysis

Each crop-climate-soil combination contributes a point to the empirical distribution function (EDF) for EEC. A variety of statistical summaries were then performed on these results: 1) the 90th percentile (1-in-10 year return concentration) was calculated for each bin within a region (e.g., Figure 4.2 on p. 132); 2) an EDF was constructed for 90th percentiles across all bins within a region (e.g., Figure 4.3 on p. 133); 3) the 90th percentile was calculated from the resulting EDF after applying area weighting factors, described as the "total acreage of suitable land for a crop within a *bin* [i.e., runoff/erosion category] relative to total acreage suitable for that crop in the region"; and 4) an EDF was constructed with the 36 annual concentrations for all bins within a region and the 90th percentile was calculated after applying the area weighting factors. EPA has several concerns with these approaches:

a. By including all crop/soil/climate scenarios in the analysis, there is less emphasis on the subset of conditions that may tend to present greater risks to aquatic receptors. To some extent, analysis #2 described above (i.e., assembling the EDF for 90th percentile concentrations) addresses the objective of evaluating the more high-risk environments, however the variability in concentrations reflects the temporal variability in rainfall events (climate/meteorology) and pesticide use (application date) combinations, rather than the soil conditions. A preferable approach would have been to apply analysis #1 and #2 to the a subset of the more conservative soil environments. For example, restricting the scenarios evaluated to hydrologic soils C and D.

b. Because exposure distributions were only presented graphically, EPA was not able to replicate the EDFs. The document should present sufficient information that all calculations and results can be replicated. Specifically, the document should include the weighting factors applied, along with references and calculations (if any) applied to the literature values.

c. Insufficient information is presented to evaluate the crop suitability weighting approach. While the concept of improving the representativeness of the scenarios is commendable, many questions would need to be addressed in order to provide the risk manager with a more complete understanding of the assumptions inherent in the approach: Does the crop suitability index correspond with the areas in which the crop is actually grown in a particular region? Does the index account for agricultural practices that can enhance the suitability of the area for a crop (e.g., irrigation, soil amendments)? In general, weighting approaches appear to be more consistent with a Tier 3, refined PRA, rather than a Tier 2 analysis.

Recommendations:

1. When applying weighting approaches, the effect of this methodology should be clearly demonstrated. For example, the percentiles and values of the weighted **and** un-weighted EDFs should be provided.

2. When subdividing large geographic areas (e.g., regions) into smaller areas with similar environmental conditions (soil, landscape, crop use, weather), characteristics of the subdivisions should be clearly presented. Each scenario should be carefully evaluated for plausibility by referring back to the conceptual site model.

3. Area-weighting approaches may be useful to consider since the wide range of soil/landscape/climate conditions are unlikely to be equally representative of areas where specific crops are grown. However, the assumptions associated with area-weighting approaches should be discussed in detail. Care should be taken that the value of introducing a more complex methodology is not offset by

raising more questions and greater uncertainty (i.e., a lateral move in Figure 1). In general, weighting schemes are more appropriately explored in Tier 3 analyses.

b. Interpretation of Discrepancies between Modeled and Monitored Atrazine Concentrations

In order to provide some perspective to the Tier 2 modeling results, a comparison is made to reported monitoring data. This is an essential step prior to planning what modeling assumptions and parameter values will be refined / explored in a Tier 3 assessment. The PRA offers the following conclusion:

> *"The predicted Tier 2 concentrations are considerably higher than those that have been reported in aquatic monitoring studies conducted for atrazine, which indicates that degradation process and attenuation processes within the agricultural landscape are not being addressed by these simulations (Giddings et al., 2000, p. 123; Williams et al., 2000, p. 41)."*

This type of statement is a premature conclusion. EPA acknowledges that Tier 2 assessments may not account for a variety of degradation and attenuation processes. However, without the benefit of a sensitivity analysis, it is unclear what sources of uncertainty may explain discrepancies between modeled and monitored data. Furthermore, potential limitations in the monitoring data should be presented - if modeled results exceed monitored results - is it more likely that the model tends to overestimate concentrations, or vice versa? Relevant topics to introduce on this issue include: 1) spatial variability (what is the proximity of the water bodies to the nearest fields?) 2) temporal variability (does the timing of the water sampling in the field studies capture the short window of time between rainfall and runoff into nearby farm ponds?); and 3) scale (were static ponds sampled that are relatively similar in size to the modeled environment?). It is appropriate to look to Tier 3 for refinements in quantifying physicochemical processes in order to reduce uncertainty in the exposure assessment, but it is premature to attribute the lion's share of the uncertainty to these variables.

It should be noted that the modeling output pertains to 90th percentiles, and the monitoring data presumably pertains to a variety of exposure percentiles. Unless the monitoring data was screened to represent only the 90th percentile of the data within a region relative to the modeled region, one would expect the model numbers to be higher.

2.C. Effects Characterization

The basic approach used in Tier 2 for effects characterization (development of an equation to characterize the full exposure response curve) is consistent with ECOFRAM and EPA agrees this approach is inherently valuable and informative. However, EPA considers the specific techniques used to develop exposure-response functions in the Tier 2 evaluation for atrazine to have substantial shortcomings.

a. Method for Characterizing Exposure-Response Curves

EPA has the following concerns regarding the methods used to characterize exposure-response curves:

1. While EPA agrees that the log-probit is a useful default for some types of exposure-response curves (mainly quantal or graded responses), alternative models (e.g., exponential) may be more appropriate for certain continuous endpoints (e.g., growth, chlorophyll production, etc). In addition, even when the log-probit is a reasonable default, it may not always be true that the model will yield a good fit with the data, and without some sort of description of the goodness of fit, it is not possible to judge if use of the log-probit is reasonable.

2. In two of three cases (*Oncorhynchus mykiss* and *Daphnia magna*), the slope of the log-probit model was not derived from the same study used to derive the point estimate of toxicity, but was extrapolated from some other study. The source of these extrapolated slopes was not documented, nor was there any justification that the assumed slopes were reasonable or appropriate.

3. In the case of *Daphnia magna*, the NOEC (140 μg/L) was assumed to be equivalent to the 10th percentile of the distribution, but there was no justification presented for this assumption nor was there any discussion of whether alternative percentile values might have been reasonable and how that might have altered the results.

4. In all cases, the log-probit models were defined from just 2 statistics (the point estimate of a measured or assumed percentile and the reported or assumed slope). In general, EPA considers that exposure-response models fit in this way have very low credibility, especially when one or both of the model parameters are assumed.

Recommendation:

All exposure-response models used in Tier 2 should be developed from the primary exposure-response data set, not from the very limited data available in the "one-liner" database. When the data are fit to a model, it is important to provide some information on the quality of the model fit to the data. This should include both a graphical comparison of the best fit equation with the raw data, and the statistical uncertainty bounds around the best fit model parameters (e.g., slope, intercept). In cases where the default log-probit model yields a poor fit, it may be appropriate to investigate the fit of other model types in subsequent refinements of the analysis. This is especially important in cases where the available data do not constrain the curve at exposures in the low end of the curve.

b. Characterizing Variability in Sensitivity Across Species

The Tier 2 effects characterization developed exposure-response functions only for the most sensitive animal (rainbow trout, Daphnia) and plant (duckweed) receptors identified in Tier 1. This approach is consistent with ECOFRAM and is generally acceptable, but EPA believes that it is desirable to provide information on the range of sensitivities among species beginning in Tier 2, rather than waiting until Tier 3.

Recommendation:

As noted above, major taxonomic groups should be maintained separately (data permitting). In addition (data permitting), exposure-response curves should be developed not only for the most sensitive receptor for each group from Tier 1, but also for the 5th percentile, the median (50th percentile), and the 95th percentile receptors in each group. These exposure response distributions should be summarized in tabular and graphical format, showing in a single figure the curves for the 5th, 50th, 95th and most sensitive receptor. This information provides a good foundation for a more extensive evaluation of inter-species variability in Tier 3, and helps indicate whether additional toxicity data may be needed to support a higher tier analysis.

2.D. Risk Characterization

Risk characterization in Tier 2 utilized the Joint Probability Curve (JPC) approach recommended by ECOFRAM. EPA has several specific concerns regarding the presentation of the results in the Tier 2 analysis.

a. Because exposure distributions were only presented graphically, EPA was not able to replicate the derivation of the JPCs. The document should present sufficient information that all calculations and results can be replicated.

b. There are several alternative methods for deriving JPCs, and the document is not entirely clear which method was employed. If the curve was calculated based on equal increments in "concentration space", this is considered to be less robust than a curve based on equal steps in "probability space" (even though the results may be essentially identical in most cases).

c. Not all JPCs are presented. Rather, a single summary statistic ("Total Risk", equal to the area under the curve or AUC) is presented. EPA believes that reducing a JPC to a summary statistic is a step in the wrong direction, since interpretation of the AUC is not straightforward, and in some cases JPCs with equal AUC may not be of equal ecological relevance. This basic view is consistent with text in the atrazine document, which states (page 55): "To fully interpret the significance of the risk, it would be necessary to assess the ecological relevance of any effects that are identified by going back to the JPC itself and the exposure and toxicity data from which it was prepared".

Recommendations:

1. The document should present sufficient information and data so that all JPC calculations and results can be replicated.

2. With only a few potential exceptions (e.g. exposure and toxicity curves that essentially do not overlap), all JPCs should be presented rather than just the AUCs. In addition, greater effort should be invested in assisting readers with the interpretation of the JPCs. In particular, the text should remind the reader of the nature of the specific exposure distribution and exposure-response function used to generate the JPC. The text on the bottom of page 122 and the top of page 123 is a good effort, but should be expanded to discuss what levels of risk might be occurring at other times besides the annual peak concentrations captured in the exposure distribution, and how this would affect conclusions about the frequency and magnitude of effect.

3. To the extent possible, the PRA should discuss the relationship between the results of the Tier 2 risk characterization and the overall assessment endpoints established for the risk assessment (presented on page 49 of the assessment). EPA understands that data may not always be available to draw a clear link, but a discussion of this point, including the limitations posed by a lack of a clear link, should be presented.

TIER 3

3.A. Problem Formulation

The PRA appropriately begins to focus on characterizing the variability and uncertainty in some of the model inputs for Tier 3. Assuming the data were not readily available already, this could involve conducting a more extensive literature search (as was illustrated by the PRA). In addition, both typical and maximum application rates were used. In addition, the exposure characterization included scenarios for small (first order) streams for corn and sorghum fields. The level of complexity and refinement introduced into Tier 3 is, in general, well explained and appropriate. The criteria by which the available data were excluded from the analysis was less clear (especially for the toxicity characterization). Furthermore, greater thought to representativeness of species for specific geographic areas should have been considered more carefully. Specific recommendations are offered below.

3.B. Exposure Characterization

Exposure characterization in Tier 3 is based on selected crop/soil/climate scenarios using PRZM/EXAMS simulations. Alternative choices for input variables were explored based on a more comprehensive review and analysis of the available data. Comments on approaches to specific variables are given below.

The PRA also included stream and river scenarios, based on simulations using PRZM 3.12 and RIVWQ 2.00 - this model was not available at the time of the review of the PRA, and therefore, is not considered in this set of comments.

a. Typical Atrazine Use Rates

Typical use rates were based on survey data from 1998 compiled by Doane Marketing Research, Inc. An area-weighting approach was used to estimate the distribution of application rates across a rather large geographic area (i.e., 5 regions across the U.S.). This spatial aggregation of the data appears to be inconsistent with the concept of the “bins” introduced into Tier 2, whereby the U.S. was divided into 11 regions and further subdivided into as many as 25 categories based on runoff/erosion potential. The data collected for the PRA up to and including Tier 3 suggest that information on application rates as well soil / landscape / meteorology could be co-located for specific geographic areas, at scales smaller than a region. This approach would yield more representative exposure scenarios.

Recommendation: Since the EEC is a direct multiple of the application rate, care should be taken that the application rates are representative of the exposure scenarios of interest. For regions in which survey data on applications rates are available, a smaller geographic scale should be used to develop area-weighted estimates of typical application rates for atrazine. Ideally, this scale would match the geographic scale used to describe other characteristics of the conceptual site model (e.g., soil properties, landscape, meteorology, etc.).

b. Aerobic Soil Metabolism (T½)

In Tiers 1 and 2, two (n =2) values for aerobic soil metabolism were considered based on the "core" studies. In Tier 3, a larger data base was considered. There is some question as to the number of studies that were evaluated (70 or 20). The PRA document (p. 71) indicates "..approximately 70 studies were reviewed...", while Williams et al. (2000, p. 29) states: "Half-life data for twenty aerobic soil metabolisms were compiled by Novartis from studies available from literature and unpublished reports". In addition, some discussion of the literature search strategy should also be presented, given that a cursory review of the literature yielded approximately 250 studies on aerobic soil metabolism of atrazine.

In Tier 3, 10 values (from 6 studies) were retained (Williams et al., 2000, Table 3-3, p. 92). The exclusion criteria, and their effect on the mean half-life, could be more completely described. Williams et al. (2000, p. 29) state that studies were omitted "..because of limited data points or elevated temperature for incubation". In a more detailed report by Burnett et al. (2000)[4], the following criteria are offered (p. 9):

- extremes in experimental conditions (e.g., temperature and soil moisture)
- soil was fabricated in the lab (vs. field collected)
- soil was amended with bacterium or an energy source
- study was an outdoor, field study
- analytical procedure, extraction method, and/or, detection limits did not generate acceptable results

There is no way of knowing how these criteria affected the summary statistics for the half-life.

Recommendation:

---

[4]Burnett, G., K. Balu, H. Barton, W. Chen, B. Gold, P. Hertl, D. Nelson, P. Scott, and K. Winton. Summary of Environmental Fate of Atrazine. Novartis No. 1213-99. June 23.

When exclusion criteria are applied to a data set, the affect of each criterion should be demonstrated. At a minimum, the number of study results omitted due to the particular criterion should be provided. A more informative summary should be presented that shows the effect of the criteria on the statistic(s) of interest (e.g., arithmetic mean, standard deviation, range). This information can be presented in tabular or graphical format.

Among the 10 values retained, 2 values appear to be duplicates of the same soil (TN, from Winkelmann et al., 1991) with relatively low $T_{1/2}$ values (20 and 21 days). How does the use of both values help to quantify the variability in aerobic soil metabolism of atrazine? Since the input for Tier 3 (like Tiers 1 and 2) is the 90% CI for the mean, this choice results in a lower half-life (i.e., less conservative value). This assumption may also be important as it impacts the half-life value calculated for anaerobic soil metabolism (equal to two times $T_{1/2}$ for aerobic soil metabolism).

The method used to calculate the 90% CI is based on the assumption that the half-lives are normally distributed. While this approach is recommended for data sets with multiple observations, uncertainties associated with assumptions of this statistical approach should be considered and presented. For example, for this particular data set, the Shapiro-Wilk test soundly rejects this assumption ($W = 0.658 <$ critical value = 0.842). Other methods for calculating the confidence interval for the mean may be evaluated in this case (e.g., other parametric approaches, bootstrap resampling).

Recommendation:

When the 90% CI is used as an input for an environmental fate model, the method used should be supported by sound statistics. Uncertainties associated with the statistical methodology should be explored and/or discussed, especially if an approach may tend to yield lower estimates of the EEC.

c. Soil-Water Partition Coefficient (Kd)

In Tiers 1 - 3, K_d was calculated form adsorption-desorption studies (organic carbon content and K_{oc}). For Tiers 1-2, the data were limited to 4 “core” studies; for Tier 3, an expanded database of 49 studies were considered. According to Williams et al. (2000, p. 31), a regression equation was *not* used to related K_d to soil properties because cation exchange capacity (CEC) “were not available in the USDA databases. According to Burnett et al. (2000, p. 12), the 49 studies were selected from a larger database of 75 values by excluding data on sediment and data lacking CEC values. For Tier 3, K_d was calculated from K_{oc} = 171 cc/g (the arithmetic mean). Some discussion, including a graph showing the distribution of K_{oc}’s, would have been appropriate to include in a Tier 3 analysis, given that this would be a source of uncertainty in the estimate of K_d. Similarly, for the regression equation approach used in Tier 4, histograms of the Foc, pH, and CEC, should be provided in a PRA. Examples are given in Figure 2.

There is a discrepancy in the data sets reported for K_d. Williams et al. (2000), Table 3-6, p. 95, gives data for n=49, but two entries for CEC are blank (Loam and Sand soils). In Burnett et al. (2000), Table 7a also reports the n=49, but gives a different set of values for loam and sand (entries #3 and #4). Also, Burnett et al. (2000, p. 12) propose that the data base be truncated at the high-end (99^{th} percentile), which would eliminate 2 entries. As a general principle, truncation is not recommended, unless it is used as a form of uncertainty analysis to demonstrate the affect of the extreme values on the model output. EPA questions why truncation wasn't explored for the low-end as well (e.g., < 1^{st} percentile).

Recommendation:

When values are selected from a data base, it is generally useful to present histograms along with summary statistics. Care should be taken to explain why data are excluded from a data base. In general, if there is a concern that extreme values may not be representative of an exposure scenario, an uncertainty analysis should be used to evaluate results both with and without the suspect data. Extreme values should be scrutinized at both ends of the distribution.

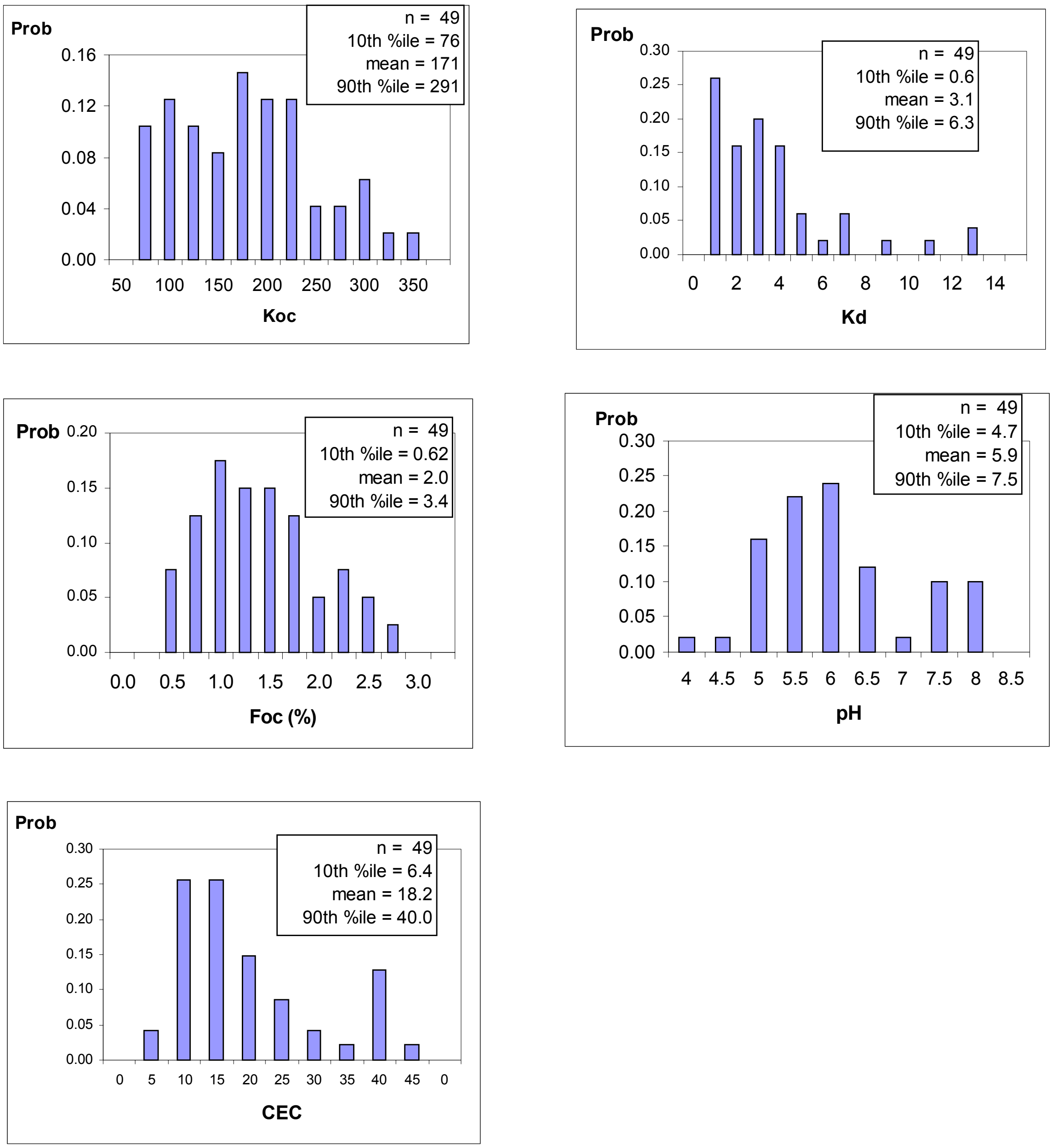

Figure 2. Examples of histograms for Tier 3 and Tier 4 assessments.

3.C. Effects Characterization

Effects characterization in Tier 3 is based on the species sensitivity distribution (SSD) approach. This is consistent with ECOFRAM, however, EPA has a number of specific comments summarized below.

a. Database Completeness

The database used to generate the SSD appears to be fairly comprehensive, but some studies with low TRVs are not included (e.g., Torres and O'Flaherty 1976).

Recommendation:

The text should provide clear rationale as to why any candidate study available in the database for a chemical is not included in the derivation of the SSD, especially if the excluded study identifies a TRV that is lower than other studies that are employed.

b. Data Grouping

The Tier 3 analysis begins to focus its exposure assessment on specific geographic sub-locations, but the effects characterization does not keep step. Toxicity data are combined into groups according to duration of exposure (acute and chronic), water type (freshwater and saltwater), and by receptor group (phytoplankton, macrophytes, zooplankton, benthos, fish, and amphibians). This grouping is reasonable, but did not take into consideration the relative likelihood that the species in each group would actually occur, or would be a suitable surrogate for species that would be expected to occur, in the geographic areas being evaluated. For example, EPA considers it unlikely that coldwater species in general and salmonids in particular will occur in significant numbers in farm ponds in the midwest. Thus, inclusion of these species might bias the SSD either too low or too high, depending on whether they are less sensitive or more sensitive that the actual receptors present in midwest farm ponds.

Recommendation:

The choice of species for inclusion in a SSD should take into consideration the suitability of the species as an indicator of the species present at the location(s) being modeled. When toxicity data are extensive, it may be appropriate to restrict the SSD to only those species that are likely to be present (or are known to be useful surrogates for species that are present). If species are included (even though they are unlikely to actually occur in the area being modeled), an analysis or rationale should be provided to justify using those data (e.g, these is no evidence that TRV values are different for those species actually present and the other species for which data are available). In addition, care should be taken to ensure that the SSD is not inappropriately weighted by multiple data for a particular sub-category of receptors.

Overall, the utility of SSDs in a Tier 3 risk assessment is open to question. Since the Tier 2 assessment presented the 5^{th}, 50^{th}, and 95^{th} percentile species results for each group of animals, how do the SSDs based on LC50, EC50 and NOEC endpoints improve our understanding of the uncertainty and magnitude of effect for each group of animals?

c. Method for Characterizing the SSD

The SSD was characterized by log-probability regression analysis. The text (Table 5.3) reported the number of data points used in the curve, the 10th percentile of the curve, and the coefficient of determination (R^2). However, the model parameters (slope, intercept) were not reported, nor was any discussion of goodness of model fit provided. Attempts to reproduce the reported statistics yielded values that were close but not always identical to those in Table 5.3.

Recommendation:

The text should provide the best fit parameters for each SSD, along with a characterization of the quality of the fit to the data. This should include a graph that shows the best fit equation compared to the raw data, as well as information on the confidence interval around the model parameters. An assumption that the species-specific TRVs are distributed lognormally is acceptable as a default, but if the quality of the lognormal fit to the data is poor, other models should be investigated.

3.D. Risk Characterization

The Tier 3 risk characterization was based on the JPC approach, except that the toxicity distributions were SSDs rather than species-specific exposure-response curves. This approach is generally consistent with ECOFRAM recommendations, but EPA has a number of specific comments on the presentation.

a. Presentation of JPCs

As noted above, EPA believes that, in general, it is not helpful to reduce a JPC to a single summary statistic (the "Total Risk", or AUC).

Recommendation:

In most cases, the full JPCs should be presented so that the shape of the distribution may be evaluated. As noted above, sufficient data should be provided so that all calculations can be reproduced.

b. Interpretation of JPCs

The PRA (p. 158) discusses the JPC results as if non-exceedence of the TRV for a species was equivalent to an absence of impact or concern for that species. For the acute SSDs, this is not correct, since the SSDs are based on acute LC50 values or EC50 values. Depending on the slope of the exposure-response curve, lethality might occur at concentrations 2-5 fold lower than the concentration that ranks as an exceedence of the TRV. In addition, how should the finding that x% of the species will be exposed to a concentration that exceeds their TRV be interpreted? How can this type of evaluation be translated into a characterization of the assessment endpoints (primary productivity, sustainability of macrophytre community structure, and long-term viability of fish populations)? If the Total Risk to animals is small, does that mean there are no animal species that are likely to be impacted?

Recommendation:

If the JPC approach is retained at the Tier 3 level, the document should provide a much more thorough discussion of the meaning and interpretation of JPCs based on SSDs, especially SSDs based on an inter-species distribution of frank effect levels (LC50, EC50). In particular, the risk characterization should make an effort to relate the findings of the Tier 3 risk characterization back to the assessment endpoints of concern.

In this regard, EPA believes that JPCs based on SSDs may sometimes be more confusing than helpful, and recommends an alternative approach be considered. In this approach, inter-species variability in effects (sensitivity) is characterized by an SSD, but risk characterization is achieved by overlaying the exposure distribution with the exposure-response curves for multiple receptors (e.g., the 5th, 25th, 50th, 75th, 90th and 95th percentile species from the SSD). This would allow ready determination of both the fraction of species that were likely to be exposed to concentrations above their respective NOEC levels, and the average severity of the response in each of those species. EPA considers this approach to be more informative than the approach utilized in the atrazine document (simply presenting the AUC for each SSD-based JPC).

TIER 4

4.A. Problem Formulation

Within the general recommendations of ECOFRAM, a Tier 4 PRA should be focused on a specific problem for which specific studies are conducted and/or analyzed to better define and characterize risk quantitatively. This may involve elaboration of dose-response or exposure assessments. ECOFRAM very explicitly indicates that a variety of different tools can be employed depending on the nature of the problem and recommends a "toolbox" rather than "cookbook" approach. Tools discussed in ECOFRAM include population level analyses, pharmacokinetic modeling, behavioral tests, microcosm/ mesocosm studies, as well as the assessment of monitoring data and watershed modeling. ECOFRAM states that, “consultation between registrants and regulators is essential at this stage because of the extraordinary cost associated with the programs.” Given the number of new approaches introduced in Tier 4 (e.g., specification of probability distributions for selected input variables, simulation of dynamic pond volumes, consideration of exposure to mixtures, etc), discussions of the conceptual approaches could have been initiated by submitting a workplan and sharing some of the newly developed models (e.g., PONDWQ).

Refined exposure scenarios were developed for two geographic locations: Tennessee and Ohio. The rationale for this selection (p. 253) is, “The systems represent high use - high rainfall and high use - moderate rainfall regions of the country, respectively.” In addition, it is clear that since one objective of the analysis in Tier 4 is to introduce monitoring data, these sites may have been selected for convenience. EPA does not disagree with the criteria, per se, but does question how these two locations were selected. Further explanation is needed as to why these two characteristics (atrazine use and rainfall) were the basis for the selection.

The Tier 4 analysis in the current PRA is largely qualitative rather than quantitative. In terms of the dose-response assessment, the analysis is limited to a qualitative assessment of various microcosm and mesocosm studies. The exposure assessment consists of a discussion of monitoring data and a Monte Carlo analysis of small ponds.

4.B. Exposure Characterization

The exposure characterization consists of a discussion of surface water monitoring, comparison of monitoring and modeling of streams, trends in exposure, mixtures of triazines and degradation products, and Monte Carlo analysis of pond scenarios. The discussion of surface water monitoring data does consist of a very clear indication of data used and data excluded (PRA, p. 233 and Table 6.5, p. 270). The decision to use a non-parametric approach (empirical rank order) because many data sets did not fit a lognormal distribution appears to be reasonable as does the discussion of biases that may be present in the available monitoring data.

The risk assessment notes (pp.243-246) that pond simulations in Tier 3 were much greater than the distribution of 15 annual maximum atrazine concentrations in Hoover Reservoir. Implicit in this discussion is the assumption that this indicates that the Tier 3 modeling was overly conservative. However, as is also (and correctly) stated in the document, small farm ponds would be expected to have higher concentrations of atrazine than larger bodies of water. The limitations inherent in comparing modeling and monitoring should be emphasized. For example, Capel and Larson (2001)[5] explore differences in atrazine loading across a wide variety of watersheds, representing different areas of use, watershed size, and rainfall patterns. Numerous sources of bias (i.e., systematic under- or overestimation of atrazine) may confound comparisons of modeled and monitored data. The PRA does raise some of these issues (p. 246). Great care should be taken, for example in comparing 90th percentiles of modeled EECs to 90th percentiles of monitoring data. As stated in the PRA, “monitoring data at a given site are incomplete reflections, at best, of the variation in concentrations over time at the site, especially for flowing water systems”. However, the PRA does not go far enough in emphasizing limitations in comparisons, as it concludes that because monitoring data are biased high (a debatable conclusion), when they fall below distributions generated by the Tier 3 models, they highlight the conservatism in the modeled estimates.

The Monte Carlo analyses of farm ponds conducted in Tier 4 were not meaningfully reviewed because programs used in the analysis (RBUFF 1.30 and PONDWQ 1.10.) were not available. The basic presentation of the modeling study (illustrated in Figure 6.17 on p. 304 of the PRA) clearly indicates that concentrations of atrazine in farm ponds are expected to be lower than concentrations in the monitored in lakes and reservoirs. This is not intuitive and is not even credible given that no attempt was made to use existing monitoring data on farm ponds to demonstrate that the modeled estimates from the Monte Carlo analysis are reasonable or even plausible. Given the large number of variables that can impact estimated concentrations in ponds (drainage area to volume, proximity to field, % crop area, tillage, soil degradation, geometry of water body), the failure to compare the pond model to monitoring data on <u>ponds</u> diminishes the credibility of the analysis.

The PRA is to be commended for addressing the temporal issues associated with atrazine use and concentrations of atrazine in ambient water (p. 249). As clearly discussed in the PRA, the use of atrazine has diminished and this reduction in use should be associated with decreased concentrations of atrazine in ambient water. The analysis of the Ohio data, however, does not indicate such a trend. While this issue is appropriately raised in the document, it remains unexplained. This suggests that our

---

[5] Capel, P.D. and S.J. Larson. 2001. Effect of scale on the behavior of atrazine in surface waters. *Environ. Sci. Tech*. 35(4): 468-657.

knowledge of the fate of atrazine in ambient water is limited. This, in turn, casts doubt on the ability to reliably model the fate of atrazine in farm ponds or any other body of water. This potentially important source of uncertainty needs to be addressed in a Tier 4 risk assessment. That the current PRA acknowledges this uncertainty is commendable. That the PRA is not able to address or reduce the uncertainty, however, suggests the conclusions based solely on modeling data (as in the pond Monte Carlo) may be unreliable.

Several of the exposure variables were quantified differently in Tier 4 than in Tier 3; the approaches used for aerobic soil metabolism ($T_{1/2}$) and Kd were addressed above (see Section 3.B) and further reviewed here.

a. *Aerobic Soil Metabolism (T½)*

In Tier 4, a regression equation is developed to relate aerobic soil metabolism to clay, pH, and CEC:

$$T_{1/2} = -385.423 + 4.388*Clay + 61.285*pH + 17.387*OC$$

The choice of these variables may be appropriate, but the relationship is overstated in Burnett et al. (2000, p. 14):

> *"Microbial population is generally expected to be dependent upon soil organic matter and clay content, which are the primary sources of nutrients for soil microbes."*

This statement is not supported. Population and population sizes or organisms responsible for atrazine metabolism may in no way relate to the overall microbial community population. Additional evidence/discussion is needed to address questions such as, "How does the microbial population depend on organic matter? What type of organic matter?" and "How does the population depend on clay?", and finally, "What are the principle organisms responsible for atrazine metabolism, and how might these vary geographically, according to different environmental conditions?".

In the Tier 4 regression equation, only 6 of the 10 values were used (Williams et al., 2000, p. 29). Four of the values were excluded because they were from foreign soils (Switzerland and Germany). The non-linear regression equation applied to this data set appears to yield an incredibly predictive model (R^2 = 0.998), almost too good to be true. However, this results is misleading. Several major sources of uncertainty should be noted: 1) the sample size is extremely small; and 2) 4 of the 6 values are clustered around 20 days, while one value exceeds 140 days. Soils with $T_{1/2}$ values between this broad range (approximately 25 - 140 days) may not be well represented at all by this analysis. A more evenly distributed data set within this range would yield a much more persuasive result for application to other sites. The PRA is to be commended for highlighting the importance of recognizing the bounds on the three exposure variables in the regression analysis.

The approach used to generate an input file for the Monte Carlo analysis by correlate values of K_d and $T_{1/2}$ to specific soil conditions is appropriate and can be very informative. In addition, a censoring approach was used to reset predicted values to the bounds (min or max) in the regression data sets whenever extrapolation extended beyond the data range used in the regression (Williams et al., 2000, p. 74). While this is a useful assumption to explore, further discussion is needed to explain the consequence of the censoring approach. What percentage of the simulated values were censored? Were values censored more at the min or max? What variable appears to determine when negative values are generated?

A sensitivity analysis should have been conducted to evaluate the importance of the input variables in both the Tennessee and Ohio simulations.

b. *Soil-Water Partition Coefficient (Kd)*

In Tier 4, a regression equation is developed to relate Kd to OC, pH, CEC:

$$Kd = 0.644 + 3.35*OC - 0.471*pH*OC + 0.0331*CEC*pH$$

Use of a non-linear regression equation is a sound approach for this analysis. See comments on $T_{1/2}$ regarding recommendations for further discussion of the uncertainties in applying the equation to two scenarios.

4.C. Effects Characterization

The effects characterization given in the Tier 4 risk assessment focuses on a qualitative discussion of mesocosm and microcosm studies, with particular emphasis on phytoplankton, periphyton, macrophytes, recovery of plant communities, changes in plant community composition, effects on invertebrates and fish.

c. Selective Use of Data

The quantitative analysis of microcosm and mesocosm studies can be extremely useful in a Tier 4 risk assessment in that it gives the risk assessor the opportunity to either confirm that lower tier dose response assessments of single species can be used to explain effects seen in more complex biological systems (validation) or to assert that standard bioassays do not appear to adequately characterize risks in more complex systems – i.e., standard bioassays may over-estimate or under-estimate risk. The current PRA, however, does not make a serious effort to quantitatively analyze microcosm and mesocosm studies or to quantitatively compare these studies to the results of lower tier bioassays.

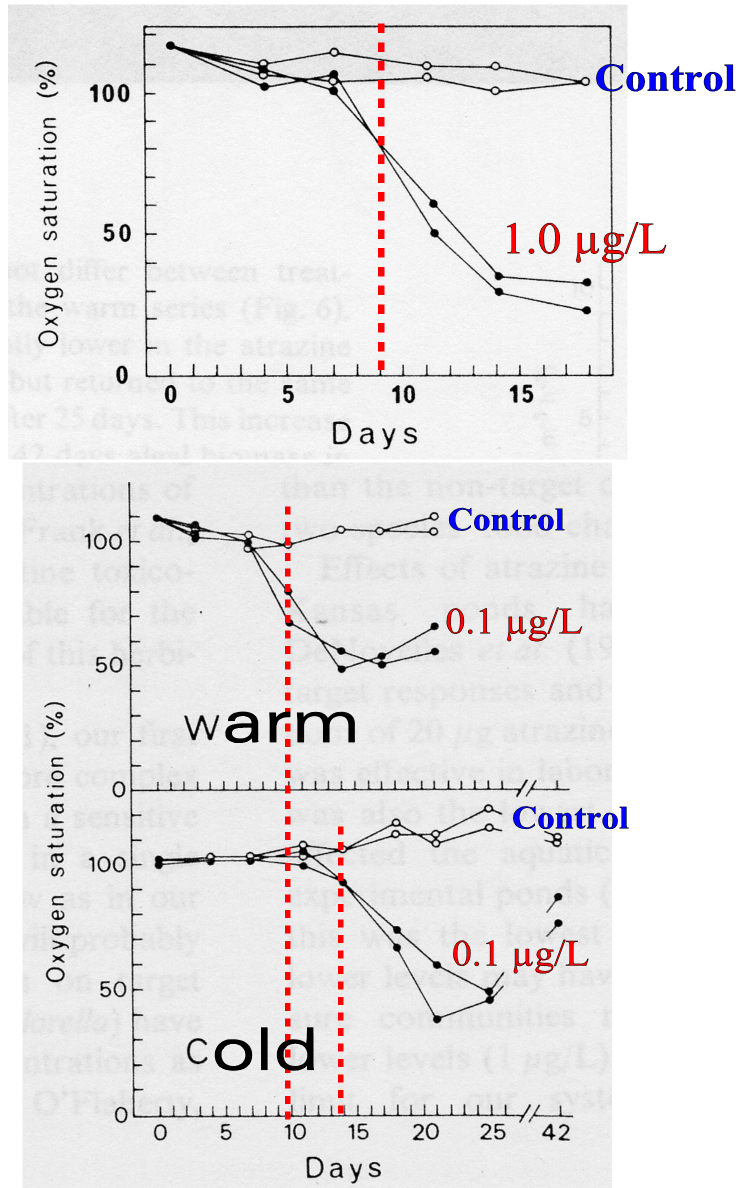

Figure 3. Effects of Atrazine on Oxygen Saturation (taken from Figures 4 and 8 in Lampert et al. 1989).

Instead, the PRA appears to dismiss or discount as unreliable a number of microcosm and mesocosm studies that report effects at low concentrations. The manner in which this is done does not appear to be analytical or objective. While not all studies cited in the PRA were examined in detail as part of this review, an examination of several studies discounted by or not fully discussed in the PRA suggests that the authors may need to re-examine some of their conclusions and expand various aspects of the discussion.

Lampert et al. 1989 - This is an early microcosm study that illustrates the utility of comparing standard bioassay data to more complex biological systems using a relatively simple system involving daphids and algae. Lampert et al. (1989) report reduced photosynthesis and chlorophyll at atrazine concentrations in water as low as 0.1 ppb. The PRA offers the following criticism:

> *"… unlike virtually all other mesocosm and microcosm studies reviewed, the effects in these enclosures did not appear until at least a week…conclusions cannot be evaluated… effects at 0.1 mg/L must be considered extremely questionable."*

A summary of Figures 4 and 8 from Lampert et al. (1989) is given in Figure 3 of this report. As illustrated in Figure 3, the data presented by Lampert et al. (1989) suggest both a dose-response and time-response relationship. A decrease in oxygen saturation occurs earlier at 1 ppb (about 9 days) than at 0.1 ppb (about 14 days in cold water and 10 days in warm water. As discussed by Lampert et al. (1989) the decreases in oxygen saturation appear to be associated with a direct effect on algae that resulted in a decrease in photosynthesis and an increase in bacterial populations, with the net effect being a decrease in oxygen concentrations in water. These patterns are evident at both 0.1 ppb and 1 ppb with duplicates in both in the control and the exposed groups.

The study by Lampert et al. (1989) may be criticized. For example, the study does not report the actual temperatures in the "warm" and "cold" studies at 0.1 ppb. In addition, the concentrations are well below the 5-day EC_{50} values of about 40 ppb reported in the Tier 1 assessment (p. 109 of PRA). The Lampert et al. (1989) study might be discounted as an unexplained outlier had the PRA gone back to the toxicity data on algae and conducted dose-response-duration analyses. If such analyses had indicated that no strong response-duration relationships were apparent, this might provide analytical support for discounting the study by Lampert et al. (1989). No such quantitative analyses are presented in the PRA. In addition, the PRA, in the discussion of the Lampert study, does not specify or detail any studies that might be used to dismiss the Lampert study. To suggest that the results are *unlike virtually all other mesocosm and microcosm studies* without providing a quantitative analysis is unconvincing.

Lakshminarayana et al. 1989 – This study attempted to assess the effects of atrazine on plankton in a stream receiving drainage from a field treated with atrazine. While the study authors indicate that the effects of atrazine appear to be *minimal*, they report a decreased abundance of phytoplankton at concentrations in the range of 1.9 ppb. This study is criticized in the PRA in the following terms: *"…many water quality variables different. …low abundance … expected .. since there was very little time for phytoplankton to become established… In short, the evidence for an atrazine effect was inconclusive"* (PRA, p. 216). One factor that the PRA specifically mentions is that differences in current in different parts of the stream may have been a factor but the PRA does not provide quantitative support for this assertion and the paper by Lakshminarayana et al. (1989) does not provide data on stream flow rates.

The basic criticisms of this study made in the PRA are reasonable. As with any field study, there may be many uncontrolled and unmeasured differences in various sites that may confound or limit the interpretation of the study. Similar criticism can, and often are, made of epidemiology studies in human health risk assessment. In reviewing such a study, particularly before the significance of the study is dismissed, some efforts should be made to quantitatively address the association between response and exposure.

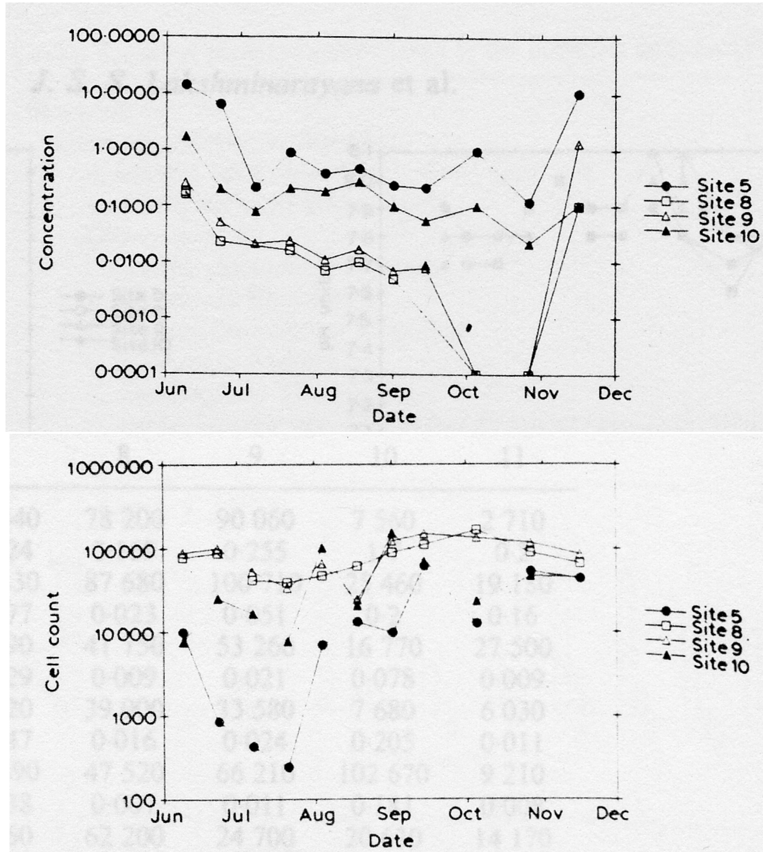

Figure 4. Measurements of atrazine concentrations in water (top graph, units of ppb) and cell counts (bottom graph, units of cells/L). (Adapted from Figures 2 and 7 in Lakshminarayana et al. 1989).

The PRA does not provide such an analysis and no such analysis is presented in Lakshminarayana et al. (1989). Nonetheless, Lakshminarayana et al. (1989) do provide sufficient data to support an at least preliminary analysis. Some of this data is presented in Figure 4 of the current report, which is adapted from Figures 2 and 7 in Lakshminarayana et al. (1989).

Figure 4 illustrates measurements of atrazine concentrations in water (top) and cell counts (bottom) taken over a period of about six months at four different sites. At Sites 8 and 9 (upstream sites), the concentrations of atrazine were generally less than 0.1 ppb and there is very little fluctuation in cell count. At Sites 5 and 10 (a channel carrying tile flow and a site down stream after mixing), the concentrations of atrazine were generally in the range of >0.1 ppb to over 10 ppb and reductions in cell count are apparent.

This visual examination of the data does not, of course, demonstrate an effect attributable to atrazine. Nonetheless, these data do provide a sufficient basis for conducting an analysis. In addition, there are other data given in the Lakshminarayana et al. (1989) [i.e., Table 3 on water quality, Table 5 on phytoplankton cell numbers, and Table 6 on phytoplankton species] and additional data (particularly on atrazine concentrations) could be available from the authors. If these data were quantitatively analyzed, an analytic and objective assessment of the Lakshminarayana et al. (1989) could be made and would be appropriate for a Tier 4 analysis.

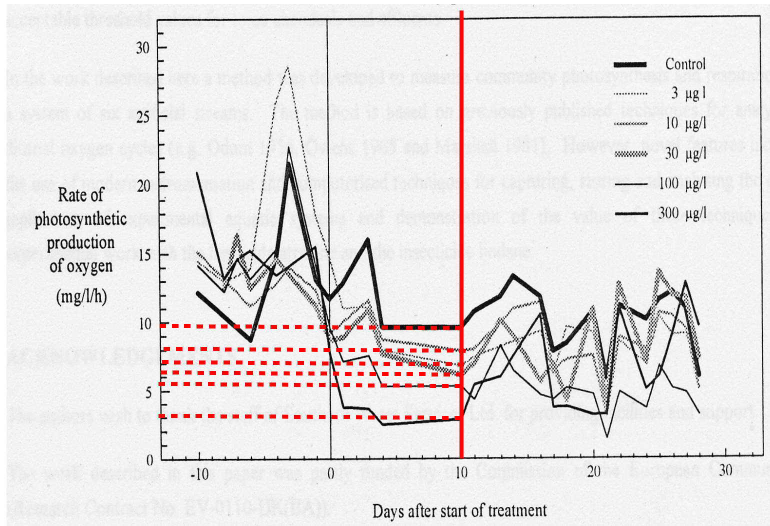

Figure 5. Effects of atrazine on rates of photosynthesis in artificial outdoor streams (taken from Figure 5 in Pearson and Crossland 1996).

Pearson and Crossland 1996 – This study attempted to assess the effects of atrazine in outdoor artificial streams containing communities of freshwater invertebrates and periphyton. Six exposure concentrations were tested (0, 3, 10, 30, 100, and 300 ppb). The effects of these exposures on rates of photosynthesis are illustrated in Figure 5 (taken directly from Figure 5 in Pearson and Crossland 1996).

The PRA presents a qualitative assessment of this study stating that oxygen production was inhibited by 100 and 300 ppb but not 3, 10, or 30 ppb. The basis for this assertion is not clear and is not supported by any quantitative analysis. There is obviously a rather variable and complex concentration-time-response relationship. In a Tier 4 analysis, it would be appropriate conduct a more quantitative analysis using the original data, if possible, or at least estimates of the measured values taken from Figure 5.

For example, there appears to be a relatively stable period of photosynthesis between Days 4 and 10. Based on estimates of the rates of photosynthesis at Day 10 (the red dashed lines in Figure 5), the apparent concentration-response relationships are plotted in Figure 6. As illustrated in Figure 6, there is a decrease in oxygen production at all concentrations (3 to 300 ppb) and this decrease appears to follow a concentration-response pattern. While Pearson and Crossland (1996, p. 918) do state that:

> *“Atrazine clearly inhibited community photosynthesis at the two highest concentrations of 100 and 300 µg/l”,*

this is different from the statement made in the PRA (p. 218):

> *"Oxygen production (indicative of periphyton photosynthesis) in outdoor recirculating stream microcosms was inhibited by 100 and 300 μg/L atrazine but not by 3, 10, or 30 μg/L (Pearson and Crossland 1996)."*

The first statement (that made by Pearson and Crossland 1996) is correct, at least as a qualitative and informal assessment in that the effects do appear to be most pronounced at 100 and 300 ppb. The statement made in the PRA, however, is unsupported and based on the Day-10 data summarized in Figure 6 the statement appears to be incorrect. Pearson and Crossland (1996) do not state that 30 μg/L is an NOEC and the data presented by Pearson and Crossland (1996) do not appear to support this assertion.

Other than some incidental summaries in tables, the above quotation from the PRA is the only information in the PRA on the Pearson and Crossland (1996) study. Additional analysis is necessary to support a Tier 4 PRA or any risk assessment. As illustrated in Figure 5, Pearson and Crossland (1996) provide sufficient data to conduct a more quantitative concentration-response assessment. Better still, a Tier 4 assessment would at least attempt to obtain the original data and then conduct a more quantitative assessment. EPA believes that this study appears to be a generally well conducted study. The inclusion of a mis-summary of the study results in the PRA is unfortunate and tends to diminish its credibility.

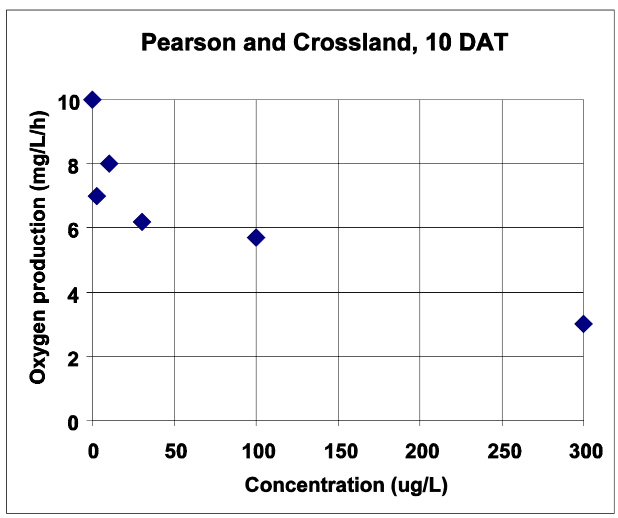

Figure 6. Apparent concentration response pattern at Day 10 (data estimated from Figure 4).

Kettle et al. 1987 – This study assessed the effects of atrazine on fish populations in artificial ponds. Six ponds were studied: 2 controls, 2 treated at 20 ppb, and 2 treated at 500 ppb. In Table 1 of the study (Kettle et al. 1987, p. 49), the number of bluegill young produced is given as 1507 and 1244 for the two control ponds, 118 and 0 for the 20 ppb ponds, and 43 and 62 for the 500 ppb ponds. As discussed by Kettle et al. 1987 (p. 51), the study was not designed to determine whether or not the observed effects were direct (toxicity) or indirect but other data presented on the stomach contents of the adult fish suggested that the effects were probably secondary to effects on macrophyte communities. This may be and is probably related to the reduced numbers of food items and prey taxa in the stomach of fish from the treated ponds that is also summarized in Table 1 of the study (Kettle et al. 1987, p. 49).

The PRA comments on this study as follows:

"The reduction in bluegill in the 20 mg/L mesocosm (which was not repeated in later years at the same concentration) is an unexplained anomaly." PRA, p. 225 to 226

The PRA does not elaborate on the statement that the results of this study was not repeated in later years. This appears to be a reference to deNoyelles et al. (1989), a study that was conducted at the same facility as that of Kettel et al. (1987) and that included Kettle as a coauthor.

The PRA characterizes the deNoyelles et al. (1989) study as finding that: *"Bluegill biomass in ponds treated with 100 to 500 µg/L atrazine was reduced by 50 to 80%, as compared with controls and ponds treated with 20 µg/L (deNoyelles et al. 1989)"* (PRA, p. 225).

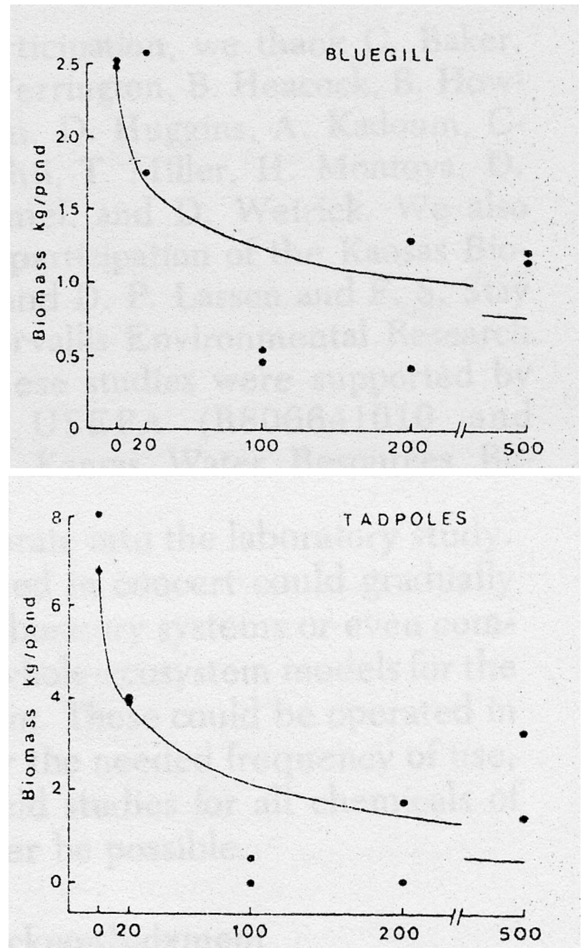

Figure 7. Relationship of atrazine concentrations to biomass of bluegills and tadpoles in artificial ponds (from Figure 7 in de Noyelles et al. 1989).

The implication that no effect on bluegill biomass was found at 20 ppb is not precise. This is illustrated in Figure 7, which is adapted from Figure 7 of de Noyelles et al. (1989). As illustrated in this figure, two ponds were tested as in the earlier study by Kettle et al. (1987). In one pond, no effect on biomass is apparent. In the other pond, biomass appears to have been reduced by about 25%.

It is not clear on what basis the PRA asserts that 20 ppb was a NOEC in the deNoyelles et al. (1989) study. If the two control ponds are simply compared to the two ponds treated at 20 ppb using a standard t-test, the differences would not be significant and 20 ppb could be characterized as an NOEC. Given the complexities and variability of artificial pond systems as well as the earlier results of Kettel et al. (1987), the weight-of-evidence suggests that 20 ppb may be at or near the threshold for observable responses and that clear responses were seen in 3 of 4 ponds.

More germaine to the criticism of the Kettel et al. (1987) study made in the PRA is the fact the deNoyelles et al. (1989) study is not a replicate of the Kettel et al. (1987) study. The deNoyelles et al. (1989) study reports data a biomass per pond and the Kettel et al. (1987) study reports data as number of young per pound. Thus, if the PRA is referring to the deNoyelles et al. (1989) study as the basis for asserting that the Kettel et al. (1987) results were not replicated, the comparison is inappropriate.

Figure 7 also includes data from the deNoyelles et al. (1989) study on amphibians. This part of the deNoyelles et al. (1989) study is not addressed in the PRA. As noted in Figure 7, there is a clear concentration-response relationship for a reduction in tadpole biomass with effects at 20 ppb. As summarized elsewhere in the PRA as part of the Tier 3 risk assessment (p. 140), LC_{50} values for amphibians have been reported as low as 30 ppb. This, while the effects on bluegills could be secondary, the reduction in populations of tadpoles could be primary (toxicity). Given concerns for amphibian populations, this sort of sensitivity could be more fully explored in a Tier 4 assessment.

Most important to the concept of a Tier 4 PRA, however, is the fact that the studies by Kettel et al. (1987) and deNoyelles et al. (1989) raise a reasonable concern that low concentrations of atrazine may impact fish and perhaps amphibian populations. Both studies can be criticized and there are uncertainties in how these studies ought to be interpreted. Nonetheless, this is precisely the kind of issue that a Tier 4 PRA should address by attempting to quantitatively consider the plausibility that the observed effects are consistent with the available information on direct action (toxicity) or secondary effects. Such an analysis would not be simple and would itself entail the quantitative assessment of both variability and uncertainty. This is what a Tier 4 assessment is intended to do in a PRA. Simply dismissing these studies with a relatively cursory qualitative discussion is unfortunate and also tends to diminish the credibility of the PRA.

4.D. Risk Characterization

Given the limitations in both the exposure assessments and dose-response assessments, comments on the risk characterization are somewhat moot. As noted in Section 4.A and detailed in the previous sections, the *Tier 4* analysis is not probabilistic and, for that matter, it is not very quantitative.

The summary of the risk characterization in the PRA (p. 16) states:

> *The question of ecological significance of acute or chronic effects on a particular fraction of species in a community was addressed by the analysis of more than 30 microcosm and mesocosm studies with atrazine. These studies demonstrated that little ecological damage and no lasting effects result from exposure of aquatic communities to atrazine concentrations of* ***50 µg/L*** *or less (the No Observed Effect Concentration for the community, $NOEC_{community}$). Effects on sensitive plant species may occur at these concentrations, but the effects are transient due to the reversibility of atrazine effects on individual plants, the resilience of algal populations, and the ecological redundancy of natural communities. The Panel did not use this information directly in the quantitative evaluation of atrazine risk, but the microcosm and mesocosm results provided a perspective on the magnitude of effects that are incorporated in the Joint Probability Curves and Total Risk estimates.*

Given the studies detailed above (as well as other studies discussed in the report), it is not clear how the risk assessment can regard 50 μg/L as a NOEC for a community. This may be related to the assessment of "recovery", as detailed in the PRA (p. 223):

> *Although the details of the taxonomic shifts may vary with the situation, the tendency for resistant species to expand into the niches vacated by sensitive species is predictable. This is perhaps the most important mechanism for ecosystem recovery from atrazine effects. In terms of total biomass and productivity, an aquatic plant community is less sensitive than its most sensitive species. "*

In other words, as long as biomass is maintained, the community is considered *recovered*. This interpretation goes well beyond the bounds of PRA but is a very important value judgment and reasonable people may disagree. A fuller discussion of the issues of community structure, diversity, and stability would be appropriate at the very least.

4.D. Recommendations

The Tier 4 section of the PRA is, for the most part, disappointing. The only "probabilistic" component of this tier is extremely difficult to review because it is not transparent. The conclusions of this portion of the Tier 4 assessment are inconsistent with any reasonable expectations and, given the failure to demonstrate that the pond model reflects any available monitoring data, the conclusions from the Monte Carlo analysis are not credible as presented.

The effects assessment in this tier is largely qualitative. The assessments of the studies that were reviewed as part of the current effort appear to be superficial and in some cases misleading. Any risk assessment must present an objective and analytical assessment of the available information. The Tier 4 risk assessment tends to rely on qualitative and subjective pronouncements rather than quantitative and objective analyses.